Impaired social learning predicts reduced real-life motivation in individuals with depression: a computational fMRI study

Article

Accepted Version

Impaired Social Learning Predicts Reduced Real-life Motivation in Individuals with Depression: A Computational fMRI Study

Anna-Lena Frey[a], Ciara McCabe[a]

[a]School of Psychology and Clinical Language Sciences, University of Reading, Reading, RG6 6AL, UK

Corresponding author

Dr Ciara McCabe
Associate Professor of Neuroscience,
School of Psychology and Clinical Language Sciences,
University of Reading,
Reading RG6 6AL,
Tel: +44 118 378 5450
c.mccabe@reading.ac.uk.

Abstract

Background: Major depressive disorder is associated with altered social functioning and impaired learning, on both the behavioural and the neural level. These deficits are likely related, considering that successful social interactions require learning to predict other people's emotional responses. Yet, there is little research examining this relation.

Methods: Forty-three individuals with high (HD; N=21) and low (LD; N=22) depression scores answered questions regarding their real-life social experiences and performed a social learning task during fMRI scanning. As part of the task, subjects learned associations between name cues and rewarding (happy faces) or aversive (fearful faces) social outcomes. Using computational modelling, behavioural and neural correlates of social learning were examined and related to real-life social experiences.

Results: HD participants reported reduced motivation to engage in real-life social activities and demonstrated elevated uncertainty about social outcomes in the task. Moreover, HD subjects displayed altered encoding of social reward predictions in the insula, temporal lobe and parietal lobe. Interestingly, across all subjects, higher task uncertainty and reduced parietal prediction encoding were associated with decreased motivation to engage in real-life social activities.

Limitations: The size of the included sample was relatively small. The results should thus be regarded as preliminary and replications in larger samples are called for.

Conclusion: Taken together, our findings suggest that reduced learning from social outcomes may impair depressed individuals' ability to predict other people's responses in real life, which renders social situations uncertain. This uncertainty, in turn, may contribute to reduced social engagement (motivation) in depression.

Keywords: depression; social; faces; learning; fMRI; computational modelling

1 Introduction

Deficits in social functioning are commonly observed in major depressive disorder (MDD; Katz, Conway, Hammen, Brennan, & Najman, 2011; Rhebergen et al., 2010; Rottenberg & Gotlib, 2008). Compared to controls, depressed individuals have fewer friends (Brim et al., 1982; Frey et al., 2019; Youngren and Lewinsohn, 1980), fewer intimate relationships (Gotlib and Lee, 1989), and spend less time with people in their social circle (Youngren and Lewinsohn, 1980). Additionally, depressed subjects show inappropriate behaviour during social interactions (reviewed in Rottenberg & Gotlib, 2008; Segrin, 2000), which can result in the receipt of negative feedback from other people (Segrin and Abramson, 1994).

Successful interpersonal interactions require learning to predict other people's responses and adjusting one's own behaviour accordingly. Therefore, social functioning abnormalities in MDD may partly be linked to impaired learning from interpersonal outcomes. In line with this suggestion, we previously found that subjects with depression symptoms show deficits in learning from social feedback and demonstrate heightened negative feedback expectancy biases during a social decision-making task. Interestingly, impaired learning predicted the experience of more negatively perceived social encounters in real life, while negative biases, as well as social anhedonia, were associated with decreased amounts of time spent with friends (Frey et al., 2019). Moreover, using a social conditioning paradigm it has previously been observed that elevated depression scores are correlated with heightened arousal ratings in response to faces that had been paired with negative statements about the participant. This effect was still seen three months after the conditioning phase, indicating that the learning of negative social associations may be enhanced in individuals with higher levels of depressive symptomatology (Wiggert et al., 2017).

The above research provides limited evidence for changes in social learning in depressed individuals. Additionally, a range of studies have reported alterations in *non-social* learning in MDD. For instance, using decision-making tasks, it has been observed that depressed

subjects display impaired reward learning (Blanco et al., 2013; Cooper et al., 2014; Herzallah et al., 2013; Kumar et al., 2018; Kunisato et al., 2012; Maddox et al., 2012; Pechtel et al., 2013; Robinson et al., 2012), while their punishment learning is either enhanced (Beevers et al., 2013; Maddox et al., 2012) or unchanged (Herzallah et al., 2013; Kumar et al., 2018; Kunisato et al., 2012; Robinson et al., 2012), when compared to controls. Moreover, in Pavlovian conditioning paradigms, depressed participants tend to demonstrate less accurate reward contingency predictions during or after the conditioning phase (Kumar et al., 2008; Robinson et al., 2012, although see Lawson et al., 2017 and Rupprechter, Stankevicius, Huys, Steele, & Seriès, 2018 for no group differences). By contrast, behavioural punishment conditioning does not seem to differ between depressed and control subjects when assessed with explicit measures (although neural group effects have been observed, see below; Lawson et al., 2017; Robinson et al., 2012).

The above behavioural research has been extended by neuroimaging studies which have examined neural learning signals with the use of computational models. In these models, the predictive value of a given cue is iteratively updated based on the difference between current outcomes and previous predictions. The latter difference, referred to as a prediction error (PE), as well as model-derived prediction values, have been used as parametric modulators in fMRI analyses (as well as to explain neural firing patterns in animal studies).

A range of brain areas have been implicated in the above learning processes (e.g. reviewed in Ernst & Paulus, 2005; Khani & Rainer, 2016; Lee, Seo, & Jung, 2012). Specifically, a network of regions including the striatum, amygdala, insula, orbitofrontal cortex (OFC) and anterior cingulate cortex (ACC) is thought to be involved in the representation of prediction values during cue presentation. In this network, the subcortical regions provide value representations which are integrated with other information, such as uncertainty and effort or delay costs, in the OFC and ACC (Bezzina et al., 2008; Croxson et al., 2009; Holland and Gallagher, 2004; Palminteri et al., 2012; Rushworth and Behrens, 2008).

Moreover, the prediction error signal is thought to be computed in the midbrain, with the substantia nigra and ventral tegmental area (VTA) representing reward PEs and the habenula encoding punishment PEs (Bromberg-Martin, Matsumoto, & Hikosaka, 2010; Cohen, Haesler, Vong, Lowell, & Uchida, 2012; Schultz, Dayan, & Montague, 1997). This PE signal is passed on to the hippocampus and striatum, where it is involved in memory acquisition and updating (Fernández et al., 2016) and value computation and action selection, respectively (Chase, Kumar, Eickhoff, & Dombrovski, 2015; Frank, 2006; O'Doherty et al., 2004).

In depressed individuals ~~display~~ reduced reward

PE encoding has been observed in the midbrain, striatum, medial orbitofrontal cortex, dorsal anterior cingulate cortex, and hippocampus, compared to controls (Gradin et al., 2011; Kumar et al., 2018, 2008; Rothkirch et al., 2017). Interestingly, the magnitude of the striatal reward PE signal has been shown to moderate the relationship between real-life anticipatory and consummatory pleasure in depressed subjects (Bakker et al., 2018). Moreover, while some studies have observed attenuated habenula punishment PE representations in depression (Liu et al., 2017), others have found these representations to be unchanged in MDD (Rothkirch et al., 2017).

In addition, examinations of neural prediction encoding have found that depressed subjects display reduced reward prediction-related responses in the hippocampus and parahippocampus (Gradin et al., 2011), as well as decreased inverse correlations between reward prediction and PE signals in the ventral striatum (Greenberg et al., 2015), compared to controls. Additionally, depressed patients demonstrate reduced punishment prediction encoding in the habenula (when shocks are used as outcomes; Lawson et al., 2017).

The above findings suggest that depression is associated with learning deficits, both on the behavioural and the neural level, partly due to impaired generation and updating of outcome predictions. However, it should be noted that most previous studies assessing learning in MDD utilised non-social outcomes. Given the ubiquity of social stimuli in everyday life, it is important to further examine how far depressed subjects' learning impairments extend to the social

domain, and whether these impairments are related to the abovementioned social functioning deficits in MDD. The current study aimed to address this question. For this purpose, a social learning task was developed in which name cues were presented followed by faces that probabilistically displayed happy, neutral, or fearful expressions. Participants with high and low depression scores completed the task during fMRI scanning and were asked to learn the average likelihood of seeing a particular emotional expression after a given name cue. Additionally, subjects answered a number of questions about their real-life social experiences. A computational model was applied to the learning task data and model-derived prediction and PE values were used as parametric modulators in the fMRI analysis to assess the neural correlates of social learning. It was hypothesised that individuals with high depression scores would show impairments in the behavioural and neural prediction of social outcomes and that these deficits would be related to deficits in real-life social experiences.

2 Methods

2.1 Participants

The current study included 43 right-handed volunteers between the age of 18 and 45 years who scored below 8 (LD; N = 21, score range: 0 to 7) or above 16 (HD; N = 22, score range: 17 to 47) on the Beck Depression Inventory (BDI, Beck, Steer, & Brown, 1996). Sample sizes were based on previous fMRI studies which detected significant group effects in (non-social) learning paradigms with 15 participants per group (Gradin et al., 2011; Kumar et al., 2008; Robinson et al., 2012). Volunteers were recruited from both the student population and the general public using flyers and posters. Subjects were screened using the structured clinical interview for DSM-IV (SCID; First, Spitzer, Gibbon, & Williams, 1996). Given that the current study was focused more generally on individuals with depression symptoms, rather than specifically on those with clinical levels of MDD, the SCID was not used for diagnostic purposes, but merely to determine if any exclusion criteria were met. Specifically, LD volunteers were excluded if they had a history of any Axis I disorder or had ever taken any psychiatric medication. HD subjects were ineligible if they had ever experienced any Axis I disorder, apart from depression and moderate levels of secondary anxiety symptoms, or if they had taken any psychiatric medication in the past year. Additional exclusion criteria for volunteers in either group were the current use of any medications besides contraceptives, the use of recreational drugs in the past three months, smoking more than five cigarettes per week, or demonstrating contraindications to MRI scanning.

The study received ethical approval from the University of Reading Ethics Committee (UREC-16/08) and was carried out in accordance with the Declaration of Helsinki. All subjects provided informed consent.

2.2 Procedure

Before the testing session, potential participants attended a screening visit during which the SCID, as well as an interview about past and current medical conditions, were conducted to ascertain that none of the exclusion criteria were met. Subsequently, a testing session was

scheduled with eligible subjects and, three days before this session, participants were sent ~~completed~~ the following online questionnaires to complete at home: trait subscale of the State and Trait Anxiety Inventory (STAI; Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983), Revised Social Anhedonia Scale (RSAS, Eckblad, Chapman, Chapman, & Mishlove, 1982), Uncertainty Intolerance Scale (UIS, Buhr & Dugas, 2002), and a demographics form, as well as the BDI (to ensure scores had stayed relatively stable; these are the reported BDI scores).

In addition, subjects answered several questions about their everyday social interactions Specifically, participants were asked 'How many friends do you have?' and 'How close do you feel to these friends' (with the latter question being rated from 1 = 'not close at all' to 10 = 'very close'). Subjects were also asked to rate the following statements (from 1 = 'strongly disagree' to 10 = 'strongly agree'): I find it difficult to make new friends; I usually really look forward to pleasant social (/non-social) activities; I usually really enjoy pleasant social (/non-social) activities; I am usually very motivated to engage in pleasant social (/non-social) activities. Ratings for social and non-social activities were made separately and, for clarification, examples of relevant activities were provided (social = meeting friends or family, dating, going to parties, etc.; non-social = going running alone, cooking alone, reading, going to museums alone, etc.).

After the above questionnaires had been completed, a testing session was arranged. At the beginning of the session, participants filled in the Positive and Negative Affect Scale (PANAS; Watson, Clark, & Tellegen, 1988). Subsequently, they performed a name learning test (see supplement) and some practice trials of the social learning task outside the MRI scanner. Following the practice, subjects completed the social learning task in the MRI scanner, and, after the scan, filled in a task feedback questionnaire.

2.3 Social Learning Task

During the social learning task, participants' aim was to learn how likely it is that a given name cue is followed by a happy, neutral or fearful facial expression. At the beginning of each trial, subjects saw one of the six names (1000ms), followed by a visual analogue rating scale (5000ms; see below). Subsequently, the face associated with the name was displayed (1000ms), showing either a neutral or an emotional expression, as determined by the probabilistic contingencies described below. The stimulus presentation was separated by a 2000ms inter-stimulus interval, and the inter-trial interval was jittered by drawing from an exponential distribution with a minimum of 2000ms and a mean of 2500ms (see Figure 1).

[Insert Figure 1 here]

The task was divided into social reward and social aversion blocks which were performed in counterbalanced order. In the social reward block, three of the six faces were displayed, each of which had a different likelihood (25%, 50% or 75%) of showing a happy rather than a neutral expression. In the social aversion block, the other three faces were presented, each of which had a different likelihood (25%, 50% or 75%) of displaying a fearful rather than a neutral expression. The six faces were randomly assigned to the blocks and likelihoods for each participant and were presented in a pseudo-random order.

Subjects were asked to learn how likely it was, on average, that a given face displayed an emotional expression. They indicated this likelihood on a visual analogue scale, ranging from 0% to 100%, in response to the question 'How likely is it that [name] is [HAPPY / AFRAID]?'. Participants were instructed to start with a guess, and to subsequently base their ratings on the intuition or 'gut feeling' they derived from all the times they had seen the name-face pairing before.

The task practice consisted of 8 repetitions of each name-face pairing, resulting in 24 trials per block and 48 practice trials in total (which were performed outside the MRI scanner). The

experimental phase (which was completed inside the MRI scanner) included 12 presentations of each name-face pairing, resulting in 36 trials per block and 72 experimental trials in total.

2.4 Analysis

2.4.1 Behavioural Analysis

Normality assumptions were not met for the questionnaire or name learning data. Group differences in these measures were therefore assessed using Mann-Whitney U tests.

Social learning task performance was examined by performing a mixed-measure (group x valence x probability) ANOVA on the likelihood ratings which were averaged across practice and experimental trials.

Moreover, to examine subjects' uncertainty regarding the task outcomes, likelihood ratings were converted into uncertainty scores. For this purpose, 50 (i.e. the value indicating maximal uncertainty) was subtracted from each likelihood rating of a given participant, separately for social reward and aversion blocks. The resulting values were transformed into absolutes and then averaged across probabilities (separately for the two blocks). This yielded two scores for each subject, with lower scores indicating higher uncertainty about what outcomes to expect. To make the result interpretation more intuitive, scores were reversed by subtracting each score from the maximum value across all participants. Thus, in the below analysis high levels of uncertainty are indicated by high uncertainty scores. A mixed-measure (group x valence) ANOVA was performed on these scores.

Additionally, to relate the learning task performance to real-life measures, uncertainty scores were entered into a regression analysis. Given that the scores for social reward and aversion blocks were highly correlated ($r = 0.57$; $p < 0.001$), scores were averaged across the two blocks. The averaged uncertainty score was then mean-centred and used to predict participants' motivation to engage in real-life social activities, together with BDI, RSAS, and mean-centred UIS negativity scores (calculated based on Sexton & Douglas 2009). An

uncertainty score*UIS negativity interaction term was also included in the analysis, as it is likely that uncertainty about social outcomes primarily affects social engagement motivation when uncertainty is perceived as negative. STAI scores were not entered into the analysis, because this would have resulted in a violation of the multicollinearity assumption (Variance Inflation Factor > 10) due to a high correlation between STAI and BDI scores. This high correlation is in line with previous findings demonstrating that the STAI contains many items that map onto depression rather than specifically onto anxiety (Bados et al., 2010). However, it should be noted that STAI scores did not significantly contribute to the prediction of motivation when they were included in the regression model and BDI scores were removed.

2.4.2 Computational Modelling

A standard Rescorla-Wagner model (Rescorla and Wagner, 1972) with a free learning rate parameter (α) was applied to the data (see supplement for details). Parameters were estimated by minimising the sum of squared errors between the model prediction value (multiplied by 100) and the participants' likelihood ratings (similar to Hindi Attar, Finckh, & Büchel, 2012). Model parameter values and fits were compared between groups using Mann-Whitney U tests.

2.4.3 fMRI Analysis

Functional MRI images were acquired using a three-Tesla Siemens scanner (Siemens AG, Erlangen, Germany) and the preprocessing and analysis of the data were performed using the Statistical Parametric Mapping software (SPM12; http://www.fil.ion.ucl.ac.uk/spm; see supplement for details). A first-level GLM analysis was conducted to examine the neural encoding of social outcome predictions. For this purpose, computational model-derived prediction values were entered as parametric modulators at the time of the cue, using separate regressors for the social reward and aversion blocks. On the second level, whole-brain one-way ANOVAs were conducted for group comparisons, which are reported at a voxelwise threshold of 0.01 (uncorrected) and are family wise error (FWE) corrected at $p<0.05$ at the

cluster-level. Moreover, to relate the fMRI results to real-life measures, parameter estimates were extracted from the peak voxels of the prediction-related group contrast and were correlated with participants' reported motivation to engage in positive social activities (similar to Gradin et al., 2011).

Additionally, neural prediction error (PE) encoding was examined. PEs reflect the difference between the predicted and the actual outcome values. Therefore, bBrain responses encoding a canonical PE should, *at the time of the outcome*, covary positively with outcome values and negatively with prediction values (derived from the computational model; see supplement). As in previous studies (e.g. Chowdhury et al., 2013; Rothkirch et al., 2017; Rutledge et al., 2017), these two PE components were thus entered into the first-level analysis as separate parametric modulators at the time of the outcome (for the social reward and social aversion block). Subsequently, MarsBar (Brett et al., 2002) was used to extract average parameter estimates for the two components from a 6mm sphere around striatal coordinates that have been found to encode PEs in a previous meta-analysis (left ROI: -10 8 -6; right ROI: 10 8 -10; Chase et al., 2015). The extracted values were then compared between groups using one-way ANOVAs.

3 Results

3.1 Behavioural Results

3.1.1 Demographic and Questionnaire Measures

Mann-Whitney U tests revealed that there were no significant group differences in age (U = 219, p = 0.970). As expected, BDI (U = 0, p < 0.001), RSAS (U = 22, p < 0.001), STAI-T (U = 0, p < 0.001), UIS negativity (U = 17, p < 0.001), and PANAS Negative Affect Scale (U = 65, p < 0.001) scores were significantly higher in HD than in LD participants. Additionally, PANAS Positive Affect Scale scores were significantly lower in HD than in LD subjects (U = 349, p = 0.001; see Table 1).

[Insert Table 1 here]

3.1.2 Real-Life Social Experiences

Compared to LD subjects, HD participants indicated having significantly fewer friends (U = 320, p = 0.001), feeling less close to their friends (U = 364, p < 0.001), and finding it more difficult to form new friendships (U = 47, p < 0.001).

Moreover, HD individuals demonstrated significantly reduced motivation to engage in pleasant social activities (U = 294, p = 0.003), as well as significantly decreased anticipation (U = 316, p < 0.001) and enjoyment (U = 323, p < 0.001) of pleasant social activities, compared to LD controls. By contrast, no group differences were observed for anticipatory (U = 223, p = 0.365), motivational (U = 227, p = 0.309), or consummatory (U = 226, p = 0.322) responses to pleasant *non-social* activities.

3.1.3 Social Learning Task Performance

A mixed measure ANOVA (group x valence x probability) performed on participants' likelihood ratings revealed the expected main effect of probability ($F(2, 82) = 94.95$, $p < 0.001$), with participants rating the likelihood of seeing an emotional expression higher after cues that were more likely to be followed by an emotional face. Moreover, a main effect of valence was observed ($F(1,41) = 8.30$, $p = 0.006$) which indicated that participants rated the overall likelihood of seeing happy faces as higher than the likelihood of seeing fearful faces. Additionally, a group by probability interaction was found ($F(2,82) = 11.77$, $p < 0.001$) which was followed up as described below. All other main effects and interactions were not significant (all $F < 2.3$).

Follow-up one-way ANOVAs revealed that, compared to LD controls, HD participants' likelihood ratings were significantly *lower* on trials with a 75% chance of showing a happy ($F(1,41) = 9.12$, $p = 0.004$) or fearful ($F(1,41) = 3.98$, $p = 0.053$) expression. By contrast, HD subjects' ratings were significantly *higher* than those of controls on trials with a 25% chance of showing a happy ($F(1,41) = 9.82$, $p = 0.003$) or fearful ($F(1,41) = 10.18$, $p = 0.003$) face (see Figure 2). No group differences were found on trials with a 50% chance of displaying a happy ($F(1,41) = 0.15$, $p = 0.698$) or fearful ($F(1,41) = 0.07$, $p = 0.796$) expression.

Moreover, a mixed-measure (group x valence) ANOVA conducted on participants' uncertainty scores (which indicate the average difference between subjects' ratings and 50%; see section 2.3.1) revealed a significant main effect of group, as HD subjects tended to be more uncertain about the social task outcomes than LD controls ($F(1,41) = 3.67$, $p = 0.062$). Additionally, a significant main effect of valence was found, showing that subjects were more uncertain about aversive than about rewarding outcomes ($F(1,41) = 6.62$, $p = 0.014$). No significant interaction effect was observed ($F(1,41) = 0.160$, $p = 0.692$).

[Insert Figure 2 here]

Additionally, a multiple regression analysis revealed that task uncertainty scores (averaged across blocks), together with questionnaire measures, predicted participants' motivation to engage in pleasant social activities ($F(5, 32) = 8.57$, $p < 0.001$, $R^2 = 0.51$). Predictors significantly contributing to this relation were the main effect of UIS negativity (β = -0.55, p = 0.008), the UIS negativity * task uncertainty interaction term (β = -0.32, p = 0.015; see Figure 3), and, marginally, RSAS social anhedonia scores (β = -0.37, p = 0.061). By contrast, the main effect of task uncertainty (β = -0.21, p = 0.096) and BDI scores (β = 0.32, p = 0.149) had no significant effect. Thus, the motivation to engage in pleasant social activities was particularly reduced in individuals who were uncertain about what social outcomes to expect and who experienced uncertainty as negative.

[Insert Figure 3 here]

3.1.4 Computational Modelling

Mann-Whitney U tests on the model parameters revealed that learning rates were significantly lower in HD than in LD participants, both in the social reward (U = 351, p = 0.004) and in the social aversion (U = 355, p = 0.003) block. The model fit, as indicated by the sum of squared errors, did not differ significantly between groups in either the social reward (U = 171, p = 0.145; U = 169, p = 0.132) or aversion (U = 189, p = 0.308; U = 182, p = 0.234) block when using individual or averaged parameters (respectively).

3.2 fMRI Results

3.2.1 Neural Prediction Value Encoding

Social reward (i.e. happy expression) prediction encoding was reduced in HD, compared to LD, subjects in the superior parietal lobe/ precuneus, as well as in a cluster including the right insula, supramarginal gyrus and superior temporal lobe (see Table 2 and Figure 4). No group differences were found for social aversion (i.e. fearful expression) prediction encoding.

Across all subjects, correlation analyses revealed a significant positive correlation between participants' motivation to engage in pleasant social activities and parameter estimates extracted from the peak prediction-related group comparison voxels in the parietal lobe (r = 0.49, p = 0.002) and insula (r = 0.36, p = 0.023). This relationship remained significant for the parietal lobe (r = 0.36, p = 0.027), but not the insula (r = 0.25, p = 0.137), when BDI and task uncertainty scores were controlled for.

[Insert Figure 4 and Table 2 here]

3.2.2 Neural Prediction Error Encoding

One-way ANOVAs were conducted on the average parameter estimates extracted from a left and a right striatal ROI for the encoding of outcome and inverse prediction values (i.e. the two PE components). No significant group differences were found for either the social reward or the social aversion block (all $F < 2.9$).

4 Discussion

4.1 Uncertainty about social outcomes predicts reduced social engagement motivation

The current study examined learning from social outcomes in individuals with high (HD) and low (LD) depression symptoms, linking task performance to measures of real-life social experiences.

It was found that, in both the social reward and the social aversion block of the learning task, HD individuals *under*estimated the likelihood of being presented with emotional faces on *high* probability trials, while they *over*estimated this likelihood on *low* probability trials (when compared to LD subjects or the actual outcome contingencies; see supplement). In other words, HD subjects provided ratings close to 50% across all trial types, indicating general uncertainty about what outcomes to expect. These findings are partly consistent with previous reports of impaired reward conditioning in depression (Kumar et al., 2008; Robinson et al., 2012; see also Chen et al., 2015). Yet, it may seem somewhat surprising that HD subjects demonstrated higher uncertainty (and thus decreased learning) in the social aversion block, considering that past studies have observed *enhanced* punishment learning in depression (Beevers et al., 2013; Maddox et al., 2012). A possible explanation of this finding is that the social stimuli used in the current study may have been particularly likely to induce rumination in HD individuals, which may have interfered with the aversion learning process (Whitmer et al., 2012). Moreover, it is worth noting that, unlike previous tasks, the current paradigm required the *continuous* formation, updating and working memory maintenance of *explicit* outcome contingencies. This may have been particularly difficult for HD individuals (independent of the stimulus valence), which would explain the general learning deficit and increase in uncertainty observed in this group.

Notably, in everyday social cognition both implicit and explicit processes play a role (Frith and Frith, 2008). Thus, HD individuals' impaired ability to explicitly predict other people's responses is likely to have an effect on real-life social functioning. In line with this suggestion,

the current study found that task-based uncertainty, in interaction with the perceived negativity of uncertainty, significantly predicted participants' motivation to engage in positive social activities (even when depression scores were controlled for). That is to say, subjects who demonstrated more uncertainty about (and thus worse learning from) social outcomes in the task, and who were more averse to uncertainty in general, were less motivated to engage in pleasant social activities in real life. It is noteworthy that HD subjects demonstrated high levels of task uncertainty, regarded uncertainty as negative, and displayed reduced social engagement motivation. Taken together, these findings suggest that deficits in learning from social outcomes may contribute to social disengagement in depressed individuals. Social withdrawal, in turn, may further increase depressed subjects' uncertainty regarding social encounters by reducing their exposure to situations in which social outcome contingencies can be learned.

The current findings are consistent with previous observations of increased intolerance of uncertainty in depression (Carleton et al., 2012). Moreover, past studies have reported a link between uncertainty intolerance and depressive rumination (Yook et al., 2010), and it has been argued that uncertainty leads to behavioural inhibition when it is regarded as negative (Carleton, 2016). It may thus be the case that, in response to higher social outcome uncertainty, depressed individuals are prone to ruminate about possible negative outcomes, which reduces (/inhibits) their motivation to engage in social activities. This idea is supported by the supplementary analysis of the present study which shows that the interaction between enhanced task uncertainty and *inhibitory* uncertainty intolerance predicts reduced social engagement motivation. In addition, the above suggestion is in line with our previous findings showing that increased negative social feedback expectancies are associated with social disengagement in individuals with high depressive symptomatology (Frey et al., 2019). It would be of interest for future studies to examine whether the relation between uncertainty and social disengagement is indeed mediated by rumination-induced negative expectancies.

4.2 Neural predication of social rewards is impaired in HD subjects

Consistent with the behavioural findings, the current study found that HD individuals displayed impaired learning signals on the neural level. Specifically, compared to controls, HD participants displayed lower covariation between social reward prediction values and BOLD responses in the superior parietal lobe, as well as in a cluster extending from the insula to the supramarginal gyrus and superior temporal lobe.

Given the superior parietal lobe's involvement in attentional processing (Behrmann et al., 2004), this region may have been recruited because the repeated pairing of cues with happy expressions made the cues more salient targets for active attentional processing. Moreover, the insula, supramarginal gyrus and temporal lobe have previously been implicated in the processing (Fusar-Poli et al., 2009) and working memory maintenance (Nichols, Kao, Verfaellie, & Gabrieli, 2006) of faces. Hence, the increased engagement of these regions by cues that were more frequently paired with task-relevant happy expressions may reflect a working memory mechanism that aids the learning processes.

Based on the above, the findings of reduced social reward prediction encoding in HD individuals in the above regions could be taken to indicate a deficit in neural attention and working memory processing during learning. However, it should be noted that BOLD responses were not simply reduced in HD subjects, but were instead reversed. That is to say, rather than being close to zero, parameter estimates extracted from the peak voxels of the group contrast were significantly below zero in the HD group (and significantly above zero in the LD group; see supplement). This indicates that, in HD individuals, BOLD responses were higher the more frequently cues were associated with *neutral* faces. A possible explanation for this finding is that, due to negative processing biases, HD individuals perceived the ambiguous neutral faces as negative, especially when they were displayed amongst happy expressions. Such a negative perception may have made the neutral faces particularly salient, and may thus have led to the recruitment of attentional and working memory resources to represent and predict neutral rather than happy faces.

The above suggestion is consistent with previous behavioural observations showing that depressed individuals tend to perceive neutral expressions as negative (Bouhuys et al., 1999; Hale et al., 1998; Leppannen et al., 2004). Moreover, the increased salience of neutral faces may also have contributed to the behavioural findings of the current study. Specifically, the mismatch between task demands (of happy expression prediction) and neural processes (of neutral expressions prediction) may have given rise to the uncertainty reflected in HD participants' task ratings. Notably, a similar mechanism could play a role in real life, if automatic processing supports learning from negative social feedback and reflective processes are needed (but potentially unable) to accurately predict the positive value of engaging in social activities (along the lines of the dual process model of Beevers, 2005).

It thus seems plausible that the neural processes of HD subjects may have supported the prediction of negatively perceived neutral expressions rather than that of happy faces. Following on from this suggestion, it may have been expected that the neural response to happy vs. neutral faces would have differed between groups, due to increased (aversive) processing of neutral faces in HD participants. Yet, such a group effect was not observed. This may potentially be the case because the prediction of neutral expressions in HD subjects, after some learning had occurred, may have engaged preparatory downregulation processes resulting in similar neural responses to neutral faces in HD and LD individuals.

Interestingly, the current study further found that lower social reward prediction encoding in the parietal lobe was significantly correlated with reduced motivation to engage in positive social activities in real life, even when task uncertainty and depression scores were controlled for. Considering the abovementioned involvement of the parietal lobe in attentional processing (Behrmann et al., 2004), this may indicate that individuals who demonstrate diminished attentional processing of positive social feedback, or enhanced attentional processing of ambiguous feedback, may be less motivated to engage in social activities (although the direction of this relation cannot be determined based on the present data). This may especially be the case in HD subjects, who displayed decreased parietal prediction encoding, as well as

reduced motivation to engage in pleasant social situations. In line with this notion, we recently found that adolescents with depression symptoms displayed blunted anticipatory responses to reward in the precuneus (and insula) and showed reduced motivation (/effort) to gain rewards (Rzepa and McCabe, 2019).

Somewhat surprisingly, and contrary to previous findings (Gradin et al., 2011; Kumar et al., 2008), the current study did not observe any group differences in PE encoding in the striatum. A potential explanation for the absence of this effect is that the utilised stimuli (happy faces of strangers) may not have been rewarding enough to elicit strong striatal PE responses in LD subjects, leading to only weak group differences. This explanation is speculative and future studies are needed to assess whether more rewarding social stimuli (such as positive pictures of close friends, partners or family members) may elicit more robust striatal PE signals in LD participants, potentially resulting in significant group effects.

4.3 Limitations

It should be noted that the current study included a relatively small sample size. Therefore, the results should be regarded as preliminary and replications in larger samples are called for. Moreover, it would be advisable for future studies to assess how social learning in depression is affected when other, including more rewarding, social stimuli ~~(besides happy, neutral and fearful faces)~~ are used.

4.4 Conclusion

All in all, the results of the current study indicate that individuals with high depression scores demonstrate impaired learning from social outcomes, on both the neural and the behavioural level. Importantly, this deficit was associated with reduced motivation to engage in real-life social activities, possibly due to increased negatively-perceived uncertainty about what to expect from social encounters. These findings tentatively suggest that improving social learning may contribute to reducing social withdrawal in depression. Future studies are needed to examine this suggestion.

References

Bados, A., Gómez-Benito, J., Balaguer, G., 2010. The state-trait anxiety inventory, trait version: Does it really measure anxiety? J. Pers. Assess. https://doi.org/10.1080/00223891.2010.513295

Bakker, J.M., Goossens, L., Kumar, P., Lange, I.M.J., Michielse, S., Schruers, K., Bastiaansen, J.A., Lieverse, R., Marcelis, M., Amelsvoort, T. Van, 2018. From laboratory to life : associating brain reward processing with real-life motivated behaviour and symptoms of depression in non-help-seeking young adults. Psychol. Med. https://doi.org/10.1017/S0033291718003446

Beck, A.T., Steer, R.A., Brown, G.K., 1996. Manual for the Beck Depression Inventory-II. San Antonio, TX Psychol. Corp. https://doi.org/10.1002/(SICI)1097-0142(19991215)86:123.3.CO;2-I

Beevers, C.G., Worthy, D.A., Gorlick, M.A., Nix, B., Chotibut, T., Todd Maddox, W., 2013. Influence of depression symptoms on history-independent reward and punishment processing. Psychiatry Res. 207, 53–60. https://doi.org/10.1016/j.psychres.2012.09.054

Behrmann, M., Geng, J.J., Shomstein, S., 2004. Parietal cortex and attention. Curr. Opin. Neurobiol. https://doi.org/10.1016/j.conb.2004.03.012

Bezzina, G., Body, S., Cheung, T.H.C., Hampson, C.L., Bradshaw, C.M., Szabadi, E., Anderson, I.M., Deakin, J.F.W., 2008. Effect of disconnecting the orbital prefrontal cortex from the nucleus accumbens core on inter-temporal choice behaviour: A quantitative analysis. Behav. Brain Res. https://doi.org/10.1016/j.bbr.2008.03.041

Blanco, N.J., Otto, A.R., Maddox, W.T., Beevers, C.G., Love, B.C., 2013. The influence of depression symptoms on exploratory decision-making. Cognition 129, 563–568. https://doi.org/10.1016/j.cognition.2013.08.018

Brett, M., Jean-Luc, A., Valabregue, R., Poline, J.-B., 2002. Region of interest analysis using

an SPM toolbox. Present. 8th Int. Conf. Funct. Mapp. Hum. Brain Sendai, Japan. Available on CD-ROM in NeuroImage.

Brim, J., Witcoff, C., Wetzel, R.D., 1982. Social Network Characteristics of Hospitalized. Psychol. Rep. 50, 423–433.

Bromberg-Martin, E.S., Matsumoto, M., Hikosaka, O., 2010. Dopamine in Motivational Control: Rewarding, Aversive, and Alerting. Neuron 68, 815–834. https://doi.org/10.1016/j.neuron.2010.11.022

Buhr, K., Dugas, M.J., 2002. The intolerance of uncertainty scale: Psychometric properties of the English version. Behav. Res. Ther. https://doi.org/10.1016/S0005-7967(01)00092-4

Carleton, N.R., 2016. Into the unknown: A review and synthesis of contemporary models involving uncertainty. J. Anxiety Disord. 39, 30–43. https://doi.org/10.1016/j.janxdis.2016.02.007

Carleton, N.R., Mulvogue, M.K., Thibodeau, M.A., McCabe, R.E., Antony, M.M., Asmundson, G.J.G., 2012. Increasingly certain about uncertainty: Intolerance of uncertainty across anxiety and depression. J. Anxiety Disord. 26, 468–479. https://doi.org/10.1016/j.janxdis.2012.01.011

Chase, H.W., Kumar, P., Eickhoff, S.B., Dombrovski, A.Y., 2015. Reinforcement learning models and their neural correlates: An activation likelihood estimation meta-analysis. Cogn. Affect. Behav. Neurosci. https://doi.org/10.3758/s13415-015-0338-7

Chen, C., Takahashi, T., Nakagawa, S., Inoue, T., Kusumi, I., 2015. Reinforcement learning in depression: A review of computational research. Neurosci. Biobehav. Rev. 55, 247–267. https://doi.org/10.1016/j.neubiorev.2015.05.005

Chowdhury, R., Guitart-Masip, M., Lambert, C., Dayan, P., Huys, Q., Düzel, E., Dolan, R.J., 2013. Dopamine restores reward prediction errors in old age. Nat. Neurosci. https://doi.org/10.1038/nn.3364

Cohen, J.Y., Haesler, S., Vong, L., Lowell, B.B., Uchida, N., 2012. Neuron-type-specific signals for reward and punishment in the ventral tegmental area. Nature 482, 85–88. https://doi.org/10.1038/nature10754

Cooper, J.A., Gorlick, M.A., Denny, T., Worthy, D.A., Beevers, C.G., Todd Maddox, W., 2014. Training attention improves decision making in individuals with elevated self-reported depressive symptoms. Cogn. Affect. Behav. Neurosci. 14, 729–741. https://doi.org/10.3758/s13415-013-0220-4

Croxson, P.L., Walton, M.E., Reilly, J.X.O., Behrens, T.E.J., Rushworth, M.F.S., 2009. Effort-Based Cost – Benefit Valuation and the Human Brain. J. Neurosci. https://doi.org/10.1523/JNEUROSCI.4515-08.2009

Eckblad, M.L., Chapman, L.J., Chapman, J.P., Mishlove, M., 1982. The Revised Social Anhedonia Scale. Unpubl. test, (copies available from T.R. Kwapil, Dep. Psychol. 850 Univ. North Carolina Greensboro, Greensboro, NC).

Ernst, M., Paulus, M.P., 2005. Neurobiology of decision making: A selective review from a neurocognitive and clinical perspective. Biol. Psychiatry. https://doi.org/10.1016/j.biopsych.2005.06.004

Fernández, R.S., Boccia, M.M., Pedreira, M.E., 2016. The fate of memory: Reconsolidation and the case of Prediction Error. Neurosci. Biobehav. Rev. 68, 423–441. https://doi.org/10.1016/j.neubiorev.2016.06.004

First, M.B., Spitzer, R.L., Gibbon, M., Williams, J.B.W., 1996. Structured Clinical Interview for DSM-IV Axis I Disorders, Clinician Version (SCID-CV). American Psychiatric Press, Inc., Washington, D.C.

Frank, M.J., 2006. Hold your horses: A dynamic computational role for the subthalamic nucleus in decision making. Neural Networks 19, 1120–1136. https://doi.org/10.1016/j.neunet.2006.03.006

Frey, A.-L., Frank, M.J., McCabe, C., 2019. Social Reinforcement Learning as a Predictor of Real-Life Experiences in Individuals with High and Low Depressive Symptomatology. Manuscr. under Rev. Prepr. available PsyArXiv. https://doi.org/https://psyarxiv.com/dq64x/

Frith, C.D., Frith, U., 2008. Implicit and Explicit Processes in Social Cognition. Neuron. https://doi.org/10.1016/j.neuron.2008.10.032

Fusar-Poli, P., Placentino, A., Carletti, F., Landi, P., Allen, P., Surguladze, S., Benedetti, F., Abbamonte, M., Gasparotti, R., Barale, F., Perez, J., McGuire, P., Politi, P., 2009. Functional atlas of emotional faces processing: A voxel-based meta-analysis of 105 functional magnetic resonance imaging studies. J. Psychiatry Neurosci. https://doi.org/10.1016/S1180-4882(09)50077-7

Garrison, J., Erdeniz, B., Done, J., 2013. Prediction error in reinforcement learning: A meta-analysis of neuroimaging studies. Neurosci. Biobehav. Rev. https://doi.org/10.1016/j.neubiorev.2013.03.023

Gotlib, I.H., Lee, C.M., 1989. The Social Functioning of Depressed Patients: A Longitudinal Assessment. J. Soc. Clin. Psychol. https://doi.org/10.1521/jscp.1989.8.3.223

Gradin, V.B., Kumar, P., Waiter, G., Ahearn, T., Stickle, C., Milders, M., Reid, I., Hall, J., Steele, J.D., 2011. Expected value and prediction error abnormalities in depression and schizophrenia. Brain 134, 1751–1764. https://doi.org/10.1093/brain/awr059

Greenberg, T., Chase, H.W., Almeida, J.R., Stiffler, R., Zevallos, C.R., Aslam, H.A., Deckersbach, T., Weyandt, S., Cooper, C., Toups, M., Carmody, T., Kurian, B., Peltier, S., Adams, P., McInnis, M.G., Oquendo, M.A., McGrath, P.J., Fava, M., Weissman, M., Parsey, R., Trivedi, M.H., Phillips, M.L., 2015. Moderation of the relationship between reward expectancy and prediction error-related ventral striatal reactivity by anhedonia in unmedicated major depressive disorder: Findings from the EMBARC study. Am. J. Psychiatry 172, 881–891. https://doi.org/10.1176/appi.ajp.2015.14050594

Herzallah, M.M., Moustafa, A.A., Natsheh, J.Y., Abdellatif, S.M., Taha, M.B., Tayem, Y.I., Sehwail, M.A., Amleh, I., Petrides, G., Myers, C.E., Gluck, M.A., 2013. Learning from negative feedback in patients with major depressive disorder is attenuated by SSRI antidepressants. Front. Integr. Neurosci. 7, 1–9. https://doi.org/10.3389/fnint.2013.00067

Hindi Attar, C., Finckh, B., Büchel, C., 2012. The influence of serotonin on fear learning. PLoS One 7. https://doi.org/10.1371/journal.pone.0042397

Holland, P.C., Gallagher, M., 2004. Amygdala-frontal interactions and reward expectancy. Curr. Opin. Neurobiol. https://doi.org/10.1016/j.conb.2004.03.007

Katz, S.J., Conway, C.C., Hammen, C.L., Brennan, P.A., Najman, J.M., 2011. Childhood social withdrawal, interpersonal impairment, and young adult depression: A mediational model. J. Abnorm. Child Psychol. 39, 1227–1238. https://doi.org/10.1007/s10802-011-9537-z

Khani, A., Rainer, G., 2016. Neural and neurochemical basis of reinforcement-guided decision making. J. Neurophysiol. 116, 724–741. https://doi.org/10.1152/jn.01113.2015

Kumar, P., Goer, F., Murray, L., Dillon, D.G., Beltzer, M.L., Cohen, A.L., Brooks, N.H., Pizzagalli, D.A., 2018. Impaired reward prediction error encoding and striatal-midbrain connectivity in depression. Neuropsychopharmacology 43, 1581–1588. https://doi.org/10.1038/s41386-018-0032-x

Kumar, P., Waiter, G., Ahearn, T., Milders, M., Reid, I., Steele, J.D., 2008. Abnormal temporal difference reward-learning signals in major depression. Brain 131, 2084–2093. https://doi.org/10.1093/brain/awn136

Kunisato, Y., Okamoto, Y., Ueda, K., Onoda, K., Okada, G., Yoshimura, S., Suzuki, S.I., Samejima, K., Yamawaki, S., 2012. Effects of depression on reward-based decision making and variability of action in probabilistic learning. J. Behav. Ther. Exp. Psychiatry 43, 1088–1094. https://doi.org/10.1016/j.jbtep.2012.05.007

Lawson, R.P., Nord, C.L., Seymour, B., Thomas, D.L., Dayan, P., Pilling, S., Roiser, J.P., 2017. Disrupted habenula function in major depression. Mol. Psychiatry 22, 202–208. https://doi.org/10.1038/mp.2016.81

Lee, D., Seo, H., Jung, M.W., 2012. Neural Basis of Reinforcement Learning and Decision Making. Annu. Rev. Neurosci. 35, 287–308. https://doi.org/10.1146/annurev-neuro-062111-150512

Liu, W.H., Valton, V., Wang, L.Z., Zhu, Y.H., Roiser, J.P., 2017. Association between habenula dysfunction and motivational symptoms in unmedicated major depressive disorder. Soc. Cogn. Affect. Neurosci. 12, 1520–1533. https://doi.org/10.1093/scan/nsx074

Maddox, W.T., Gorlick, M.A., Worthy, D.A., Beevers, C.G., 2012. Depressive symptoms enhance loss-minimization, but attenuate gain-maximization in history-dependent decision-making. Cognition 125, 118–124. https://doi.org/10.1016/j.cognition.2012.06.011

Nichols, E.A., Kao, Y.C., Verfaellie, M., Gabrieli, J.D.E., 2006. Working memory and long-term memory for faces: Evidence from fMRI and global amnesia for involvement of the medial temporal lobes. Hippocampus. https://doi.org/10.1002/hipo.20190

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., Dolan, R.J., 2004. Dissociable Roles of Ventral and Dorsal Striatum in Instrumental Conditioning. Science (80-.). 304, 452–454. https://doi.org/10.1126/science.1094285

Palminteri, S., Justo, D., Jauffret, C., Pavlicek, B., Dauta, A., Delmaire, C., Czernecki, V., Karachi, C., Capelle, L., Durr, A., Pessiglione, M., 2012. Critical Roles for Anterior Insula and Dorsal Striatum in Punishment-Based Avoidance Learning. Neuron. https://doi.org/10.1016/j.neuron.2012.10.017

Pechtel, P., Dutra, S.J., Goetz, E.L., Pizzagalli, D.A., 2013. Blunted reward responsiveness in remitted depression. J. Psychiatr. Res. 47, 1864–1869.

https://doi.org/10.1016/j.jpsychires.2013.08.011

Rescorla, R., Wagner, A., 1972. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement, in: Classical Conditioning: Current Research and Theory, Vol. 2. https://doi.org/10.1101/gr.110528.110

Rhebergen, D., Beekman, A.T.F., de Graaf, R., Nolen, W.A., Spijker, J., Hoogendijk, W.J., Penninx, B.W.J.H., 2010. Trajectories of recovery of social and physical functioning in major depression, dysthymic disorder and double depression: A 3-year follow-up. J. Affect. Disord. 124, 148–156. https://doi.org/10.1016/j.jad.2009.10.029

Robinson, O.J., Cools, R., Carlisi, C.O., Sahakian, B.J., Drevets, W.C., 2012. Ventral striatum response during reward and punishment reversal learning in unmedicated major depressive disorder. Am. J. Psychiatry 169, 152–159. https://doi.org/10.1176/appi.ajp.2011.11010137

Rothkirch, M., Tonn, J., Köhler, S., Sterzer, P., 2017. Neural mechanisms of reinforcement learning in unmedicated patients with major depressive disorder. Brain 140, 1147–1157. https://doi.org/10.1093/brain/awx025

Rottenberg, J., Gotlib, I.H., 2008. Socioemotional Functioning in Depression, in: Mood Disorders: A Handbook of Science and Practice. https://doi.org/10.1002/9780470696385.ch4

Rupprechter, S., Stankevicius, A., Huys, Q.J.M., Steele, J.D., Seriès, P., 2018. Major Depression Impairs the Use of Reward Values for Decision-Making. Sci. Rep. 8, 1–8. https://doi.org/10.1038/s41598-018-31730-w

Rushworth, M.F.S., Behrens, T.E.J., 2008. Choice, uncertainty and value in prefrontal and cingulate cortex. Nat. Neurosci. https://doi.org/10.1038/nn2066

Rutledge, R.B., Moutoussis, M., Smittenaar, P., Zeidman, P., Taylor, T., Hrynkiewicz, L., Lam, J., Skandali, N., Siegel, J.Z., Ousdal, O.T., Prabhu, G., Dayan, P., Fonagy, P., Dolan,

R.J., 2017. Association of neural and emotional impacts of reward prediction errors with major depression. JAMA Psychiatry 74, 790–797. https://doi.org/10.1001/jamapsychiatry.2017.1713

Rzepa, E., McCabe, C., 2019. Dimensional Anhedonia and the Adolescent brain: Reward and Aversion Anticipation, Effort and Consummation. Manuscr. under Rev. https://doi.org/10.1101/473835

Schultz, W., Dayan, P., Montague, P.R., 1997. A neural substrate of prediction and reward. Science (80-.). https://doi.org/10.1126/science.275.5306.1593

Segrin, C., 2000. Social Skills Deficits Associated With Depression. Clin. Psychol. Rev. 20, 379–403.

Segrin, C., Abramson, L.Y., 1994. Negative Reactions to Depressive Behaviors 103, 655–668.

Spielberger, C.D., Gorsuch, R., Lushene, R., Vagg, P., Jacobs, G., 1983. Manual for the State-Trait Anxiety Inventory (STAI Form Y). Consulting Psychologists Press, Palo Alto. https://doi.org/10.5370/JEET.2014.9.2.478

Watson, D., Clark, L.A., Tellegen, A., 1988. Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales. J. Pers. Soc. Psychol. https://doi.org/10.1037/0022-3514.54.6.1063

Whitmer, A.J., Frank, M.J., Gotlib, I.H., 2012. Sensitivity to reward and punishment in major depressive disorder: Effects of rumination and of single versus multiple experiences. Cogn. Emot. 26, 1475–1485. https://doi.org/10.1080/02699931.2012.682973

Wiggert, N., Wilhelm, F.H., Boger, S., Georgii, C., Klimesch, W., Blechert, J., 2017. Social Pavlovian conditioning: Short- and long-term effects and the role of anxiety and depressive symptoms. Soc. Cogn. Affect. Neurosci. 12, 329–339. https://doi.org/10.1093/scan/nsw128

Yook, K., Kim, K.H., Suh, S.Y., Lee, K.S., 2010. Intolerance of uncertainty, worry, and rumination in major depressive disorder and generalized anxiety disorder. J. Anxiety Disord. 24, 623–628. https://doi.org/10.1016/j.janxdis.2010.04.003

Youngren, M.A., Lewinsohn, P.M., 1980. The functional relation between depression and problematic interpersonal behavior. J. Abnorm. Psychol. https://doi.org/10.1080/08977190500096004

Figure 1: *Example of a social learning task trial.*

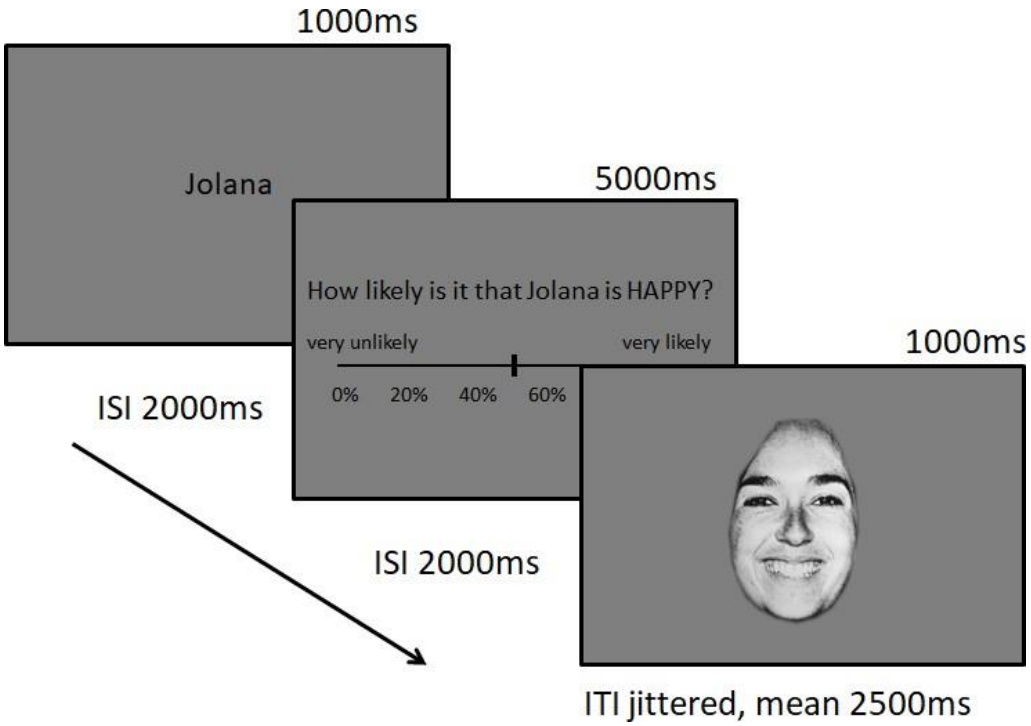

Figure 2: *Likelihood ratings by chance of seeing an emotional face for A) the social reward and B) the social aversion block in individuals with high (HD) and low (LD) depression scores.*

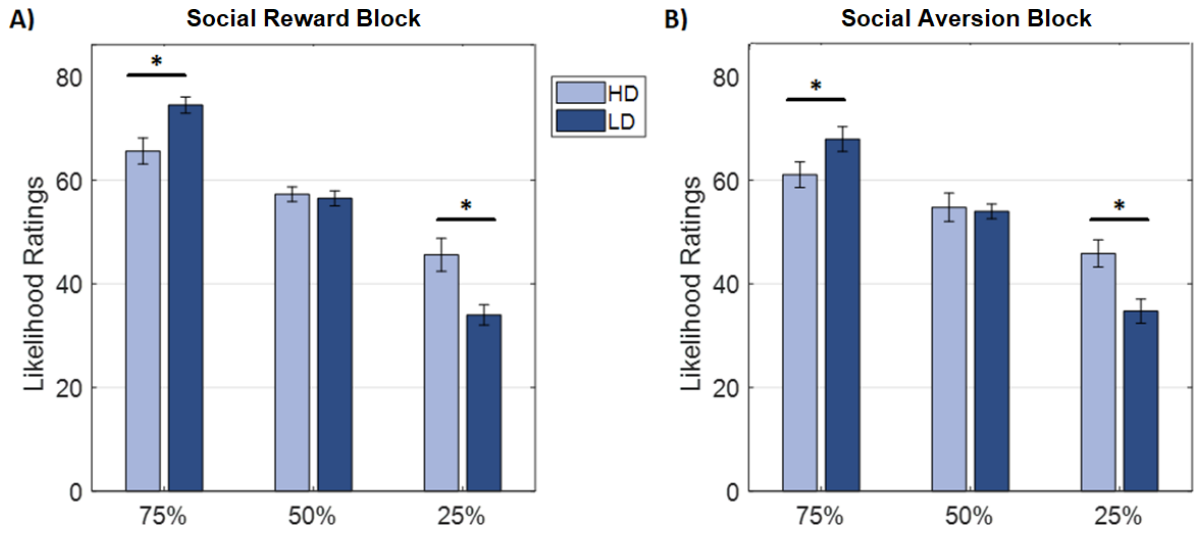

Figure 3: *Scatter plot showing the association between motivation to engage in pleasant social activities (higher scores indicate higher motivation) and uncertainty intolerance (UIS) * task uncertainty interaction values.*

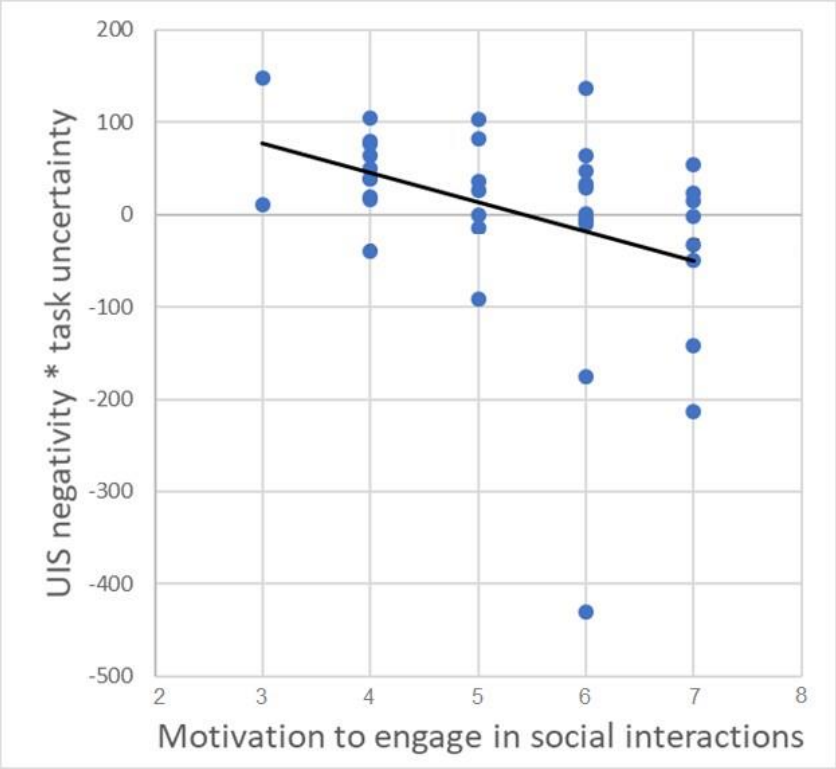

Figure 4: *Clusters showing lower social reward prediction encoding in individuals with high (HD) than with low (LD) depression scores, as well as parameter estimates extracted from A) the insula peak voxel and B) the parietal peak voxel.*

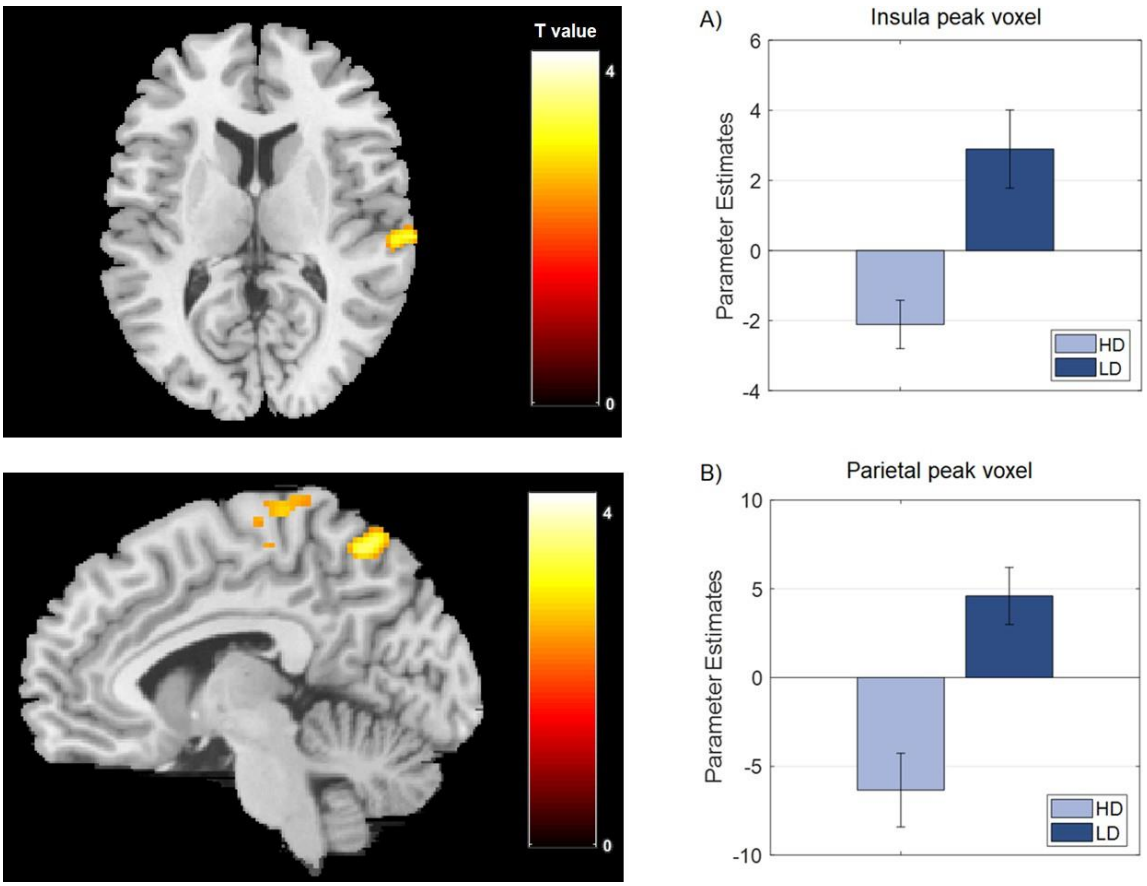

Table 1: *Demographic data and questionnaire scores for individuals with high (HD) and low (LD) depression scores.*

	HD (N = 21)		LD (N = 22)	
	Mean	SD	Mean	SD
Age (years)	23.20	5.66	22.45	4.35
N females/ males	17/4	-	14/8	-
BDI*	26.05	9.63	1.36	1.84
RSAS*	18.57	6.43	5.77	4.31
STAI-T*	57.75	7.12	27.85	6.92
UIS - neg*	94.71	17.81	52.76	17.19
PANAS - pos*	24.38	5.71	31.52	6.57
PANAS - neg*	21.29	7.27	13.43	5.26

SD, standard deviation; BDI, Beck Depression Inventory; RSAS, Revised Social Anhedonia Scale; STAS-T, trait score of the State Trait Anxiety Inventory; UIS - neg, Uncertainty Intolerance Negativity Scale; PANAS-pos/neg, positive and negative mood scores of the Positive and Negative Affect Scale; * asterisks indicate significant group differences

Table 2. *Parametric modulation results for social reward prediction encoding in individuals with low (LD) vs. high (HD) depression scores.*

	MNI coordinates				
Brain Region	X	Y	Z	Z score	*p* value
LD > HD					
Superior Parietal Lobe/ Precuneus	-18	-58	68	3.80	0.001
Right Insula	48	-20	18	3.47	0.045
Right Supramarginal Gyrus	58	-32	24	3.28	
Right Superior Temporal Lobe	68	-22	12	3.17	

Whole-brain cluster p values are family-wise error corrected at $p < .05$

Disclosures

The authors report no conflicts of interest.

Author Statement

Contributors

Anna-Lena Frey designed the study, carried out the data collection (with some assistance from Evelyn Toh and Canan Asli Can), performed the statistical analysis, and wrote the manuscript. Ciara McCabe provided support and gave feedback on the study design and on the manuscript. Anna-Lena Frey and Ciara McCabe contributed to and approved the final manuscript.

Funding

This work was supported by the Medical Research Council PhD studentship of AF.

Acknowledgements

We would like to thank Evelyn Toh and Canan Asli Can for their assistance with the data collection.

Supplement

Supplementary Methods

Name Learning Test

Before completing the social learning task, subjects were asked to rate their familiarity and their positive and negative associations with a list of modified Scandinavian and Eastern European names (on a scale from 0 = 'no association/ familiarity' to 10 = 'strong association/ familiarity'). The names with which participants were least familiar, and with which they had the weakest associations, were chosen as cues for the social learning task on an individual basis.

As described in the main paper (section 2.3), the social learning task involved learning how likely it is that a given name cue is followed by a face with a happy, neutral or fearful expression, while the face *identity* that a particular name is paired with stays constant. To ensure that participants were fully focused on learning the name-emotion associations during the task, subjects were asked to memorise the name-face identity pairings beforehand. For this purpose, participants were shown the selected names together with the (neutral) faces that were going to be used during the learning task (i.e. three male and three female faces from the Pictures of Facial Affect Series; Ekman & Friesen, 1976). Subjects were given as much time as they needed to memorise the name-face identity pairings. Once they felt ready, participants completed a name learning test, during which the six faces were numbered and displayed in a random order together with *one* of the learned names. Subjects were instructed to select the number of the face that was associated with the presented name. After each choice, the words 'correct' or 'wrong – the correct face is:' were displayed for one second together with the correct face. The name test continued until participants had correctly matched each name with the corresponding face three times. The order in which the names were displayed was pseudo-random. Participants' memorising time, accuracy, reaction times, and number of trials needed to reach criterion were recorded.

Computational Modelling

A Rescorla-Wagner model (Rescorla and Wagner, 1972) was applied to the data, in which the prediction error (δ) for a given trial (t) was calculated as the difference between the predicted value (V) and the actual outcome (r):

$$\delta = r(t) - V(t)$$

Moreover, the predicted value for the next trial was updated by adding the prediction error, multiplied by a learning rate (α), to the previous prediction:

$$V(t+1) = V(t) + \alpha * \delta$$

The predicted value was (at first, see below) initialised at 0.5, which reflects the mean probability of encountering an emotional (rather than a neutral) expression, as well as the fact that it is reasonable for participants to initially rate the likelihood of seeing an emotional expression as 50% (expressing maximal uncertainty). Moreover, outcome values were coded as 0 for neutral expressions and as 1 for happy or fearful faces, thus capturing the prediction of salient emotional outcomes. It should be noted that coding fearful faces as -1 (and initialising V at -0.5) simply leads to a change in sign of the prediction and prediction error values compared to coding fearful expressions as 1. The negative encoding of fear predictions can thus be assessed by examining negative covariations between prediction values and BOLD responses in the below parametric modulation fMRI analysis.

Given that the same stimuli and outcome contingencies were used during the practice and experimental phases of the social learning task, the computational model was fit to participants' data across both phases, but separately for social reward (happy) and aversion (fear) blocks. To account for the fact that forgetting was likely to occur between the practice and experimental trials, which were performed outside and inside the MRI scanner, respectively, all prediction values were decayed towards the initial value of 0.5 after the 48 practice trials:

$$V(49) = V(49) + \gamma * (0.5 - V(49))$$

where γ is the decay parameter determining the strength of the 'forgetting' effect. (A similar method has been used by Collins & Frank, 2012 to capture the effects of working memory decay.)

The decay and learning rate parameters were estimated for each participant by minimising the sum of squared errors between the model prediction value (V, multiplied by 100) and the participant's likelihood ratings (similar to Hindi Attar, Finckh, & Büchel, 2012). The fitting procedure was performed in two steps, because the practice data were missing for four HD and nine LD participants (due to technical difficulties). Firstly, the model was fit to only the data of those participants for whom the practice data was available. Using the estimated parameters, the prediction values (V) for the first experimental trial of each stimulus were obtained for each included participant. These prediction values were then averaged across subjects. Subsequently, the model fitting was repeated for *all* participants for *only the experimental trials* (thus estimating only α and not γ), utilising the average prediction values from the first fitting step to initialise V (instead of using 0.5). In this way, the learning that occurred during the practice trials was taken into account for all subjects, without biasing the model fitting depending on whether or not practice data was available for a given participant (as V was initialised at the same value for *all* participants). Note that, for those participants for whom experimental *and* practice data were available, the model fit and the parameter estimates were highly similar during the first and second step of the fitting procedure, indicating that this approach does not seem to negatively affect the parameter estimation.

Moreover, it is worth noting that the main purpose of the computational modelling analysis was to derive prediction and prediction error values for the parametric modulation fMRI analysis. A previous systematic exploration of the effect of model parameter values on fMRI results revealed that parametric modulation results did not differ substantially as learning rates (and therefore prediction and prediction error values) were varied (Wilson and Niv, 2015). Thus,

the variations due to the missing data are very unlikely to have had a notable effect on the fMRI results.

To assess group differences, Mann-Whitney U tests were conducted on the parameter estimates, as well as on the sum of squared error values (which provide a measure of model fit).

fMRI Data Acquisition

A three-Tesla Siemens scanner (Siemens AG, Erlangen, Germany) with a 32-channel head coil was used to acquire blood oxygenation level dependent (BOLD) functional images. A GRAPPA multiband sequence was utilised with an acceleration factor of 6, a repetition time (TR) of 700ms, an echo time (TE) of 30ms, and a flip angle (FA) of 90°. The whole brain was covered by the field of view (FOV) with a voxel resolution of 2.4 x 2.4 x 2.4mm^3. Additionally, structural T1-weighted images were obtained with a magnetisation prepared rapid acquisition gradient echo sequence (TR = 2020ms, TE = 3.02ms, FA = 9°) with a FOV covering the whole brain and a voxel resolution of 1 x 1 x 1mm^3.

fMRI Analysis

Preprocessing and analysis of the fMRI data was performed using the Statistical Parametric Mapping software (SPM12; http://www.fil.ion.ucl.ac.uk/spm). Functional images were realigned to the average position and motion parameters were saved for inclusion as regressors of no interest in the first-level analysis. Structural images were co-registered with the functional images and aligned to the SPM MNI space tissue probability map using segmentation. The resulting normalisation parameters were applied to the functional images which were subsequently smoothed with a Gaussian kernel of 6mm full-width at half-maximum.

Three first-level GLM analyses were run. GLM1 examined covariations between BOLD responses and values derived from the computational model described above. For this

purpose, model-derived prediction values were entered as parametric modulators at the time of the cue, using separate regressors for the social reward and aversion blocks. In line with the previous literature, prediction values were calculated using average learning rate parameters across all participants (social reward block: $\alpha = 0.12$, social aversion block: $\alpha = 0.08$) to ensure that any group differences in the fMRI results were not due to the use of varying parameter values (Bakker et al., 2018; Daw, 2011; Daw et al., 2006; Pessiglione et al., 2006; Schonberg et al., 2010, 2007). However, for completeness, the above analysis was also run with individual learning rate values (GLM2), which yielded very similar results (see supplementary fMRI results below).

As has been commonly reported in the previous literature (e.g. Behrens, Hunt, Woolrich, & Rushworth, 2009; Chowdhury et al., 2013; Rothkirch et al., 2017; Tobia et al., 2014), the outcome and prediction error (PE) values were highly correlated in the current study. It was, therefore, not feasible to unambiguously identify PE-related BOLD responses by using PE values as parametric modulators at the time of the outcome. Notably, brain responses encoding a canonical PE should, at the time of the outcome, covary positively with outcome values and negatively with prediction values. As in previous studies (e.g. Chowdhury et al., 2013; Rothkirch et al., 2017; Rutledge et al., 2017), these two PE components were thus entered into the analysis as separate parametric modulators at the time of the outcome. Subsequently, MarsBar (Brett, Jean-Luc, Valabregue, & Poline, 2002) was used to extract average parameter estimates for outcome and inverse prediction encoding from a 6mm sphere around striatal coordinates that have been found to encode PEs in a previous meta-analysis (left ROI: -10 8 -6; right ROI: 10 8 -10; Chase et al., 2015). The extracted values were then compared between groups using one-way ANOVAs.

Additionally, a third GLM analysis was performed (GLM3) to assess valence-dependent BOLD responses to the cues and outcomes. Onset timings of the following events were entered as regressors: name cues from the social aversion block, name cues from the social reward

block, fearful faces, happy faces, and neutral faces. Subsequently, contrasts were computed for social reward vs. aversion cues, fearful vs. neutral faces, and happy vs. neutral faces.

In all three GLM analyses, the regressors of interest, as well as their temporal derivatives, were convolved with the haemodynamic response function. Moreover, the six motion parameters from the realignment preprocessing step and a constant, as well as the onsets of the rating scale, were included as regressors of no interest.

On the second level, whole-brain one-sample t-tests were performed on the data of the LD control group to assess main effects, and whole-brain one-way ANOVAs were conducted for group comparisons. All results are reported at a voxelwise threshold of 0.01 (uncorrected) and are family wise error (FWE) corrected at $p<0.05$ at the cluster-level.

Finally, to relate the fMRI results to real-life measures, parameter estimates were extracted from the peak voxels of the prediction-related group comparison and were correlated with participants' reported motivation to engage in positive social interactions (similar to Gradin et al., 2011).

Supplementary Behavioural Results

Name Learning Test Performance

For the name learning test, Mann-Whitney U tests showed no significant group differences in the memorising time (U = 86, p = 0.320), accuracy (U = 88, p = 0.363), reaction times (U = 135, p = 0.320), or number of trials needed to reach criterion (U = 126, p = 0.536). Thus, there was no indication that HD subjects displayed any general deficits in associative learning (between names and face identities).

Social Learning Task Performance – Experimental Data Only

As mentioned above, the name test and social learning task *practice* data were lost for four HD and nine LD participants, due to technical difficulties. The mixed-measure (group x valence x probability) ANOVA reported in the main paper was performed on the likelihood ratings averaged across all available (practice and/or experimental) data for each participant. However, to ensure that the results were not biased by the missing data, the analysis was repeated using only the data from the experimental trials (which were available for all participants). The pattern of findings was almost identical for the two approaches (see section 3.1.3 in the main paper).

Specifically, using the experimental data only, a mixed measure ANOVA (group x valence x probability) performed on participants' likelihood ratings revealed the expected main effect of probability ($F(2, 82) = 82.39$, $p < 0.001$), as participants rated the likelihood of seeing an emotional expression higher after cues that were more likely to be followed by an emotional face. Moreover, a main effect of valence was observed ($F(1,41) = 4.35$, $p = 0.043$) which indicated that participants rated the overall likelihood of seeing happy faces as higher than the likelihood of seeing fearful faces. Additionally, a group by probability interaction was found ($F(2,82) = 8.46$, $p < 0.001$) which was followed up as described below. All other main effects and interactions were not significant (all $F < 2.1$).

Follow-up one-way ANOVAs revealed that, compared to LD controls, HD participants' likelihood ratings were significantly *lower* on trials with a 75% chance of showing a happy face ($F(1,41) = 7.59$, $p = 0.009$). By contrast, HD subjects' ratings were significantly *higher* than those of controls on trials with a 25% chance of showing a happy ($F(1,41) = 7.69$, $p = 0.008$) or fearful ($F(1,41) = 6.95$, $p = 0.012$) face. There were no group differences on trials with a 50% chance of showing a happy ($F(1,41) = 0.001$, $p = 0.976$) or fearful ($F(1,41) = 0.07$, $p = 0.794$) expression, nor on trials with a 75% chance of displaying a fearful face ($F(1,41) = 1.38$, $p = 0.248$).

Social Learning Task Performance – Accuracy

In order to examine whether there were group differences in the accuracy of the likelihood ratings, the absolute of the difference between participants' ratings and the true likelihood were calculated and entered into a mixed measure ANOVA (group x valence x probability). This analysis revealed a main effect of probability ($F(1.81, 74.23) = 21.29$, $p < 0.001$), as well as a main effect of group ($F(1,41) = 15.88$, $p < 0.001$), with LD participants making significantly more accurate ratings than HD subjects. In addition, group by probability ($F(1.81,74.23) = 4.86$, $p = 0.013$) and valence by probability ($F(1.88,77.25) = 3.85$, $p = 0.028$) interactions were observed, which were followed up as described below. All other main effects and interactions were not significant (all $F < 2.5$).

Follow-up one-way ANOVAs revealed that, compared to LD controls, HD participants' likelihood ratings were significantly less accurate on trials with a 75% chance of showing a happy ($F(1,41) = 10.50$, $p = 0.002$) or fearful ($F(1,41) = 4.10$, $p = 0.049$) expression, as well as on trials with a 25% chance of showing a happy ($F(1,41) = 10.37$ $p = 0.003$) or fearful ($F(1,41) = 8.15$, $p = 0.007$) expression. By contrast, no significant group differences were observed on trials with a 50% chance of showing a happy ($F(1,41) < 0.01$, $p = 0.958$) or fearful ($F(1,41) = 1.80$, $p = 0.187$) face.

Prediction of Social Engagement Motivation with Inhibitory Uncertainty Intolerance

Inhibitory uncertainty intolerance (UI) scores were significantly higher in HD than in LD participants (U = 31.5, p < 0.001; HD: M = 17.00, SD = 4.34; LD: M = 8.18, SD = 3.19). Moreover, similar results were obtained when predicting social engagement motivation using inhibitory UI than when utilising UIS negativity scores (as in section 3.1.3 in the main paper). Specifically, a multiple regression analysis revealed that task-based uncertainty scores and questionnaire measures predicted participants' motivation to engage in pleasant social activities (F(5, 33) = 9.35, p < 0.001, R^2 = 0.52). Predictors significantly contributing to the relation were the main effect of inhibitory UI (β = -0.53, p = 0.005), the inhibitory UI* task uncertainty interaction term (β = -0.32, p = 0.011), and RSAS social anhedonia scores (β = -0.40, p = 0.036). By contrast the main effect of task uncertainty (β = -0.17, p = 0.161) and BDI scores (β = 0.31, p = 0.143) had no significant effect. Thus, the motivation to engage in pleasant social activities was particularly reduced in individuals who were uncertain about what social outcomes to expect and for whom uncertainty had an inhibitory effect.

Task Feedback Questionnaire

In a task feedback questionnaire, HD subjects demonstrated a tendency to show higher emotional responses to fearful expressions than controls (U = 142, p = 0.069), while their self-rated ability to remember happy faces was marginally decreased (U = 280, p = 0.065). No group differences were found for emotional responses to happy faces (U = 229, p = 0.615), or for the reported ability to remember fearful faces (U = 245, p = 0.363).

Supplementary fMRI Results

Neural Prediction Value Encoding

Main Effects

In the LD group, a significant covariation between BOLD responses and model-based social reward (i.e. happy expression) prediction values was observed in a right-lateralised cluster ranging from the superior to the inferior temporal lobe and the fusiform gyrus (see Table S1). By contrast, no significant (positive or negative) covariation between BOLD responses and social aversion (i.e. fearful expression) prediction values was found.

Table S1: *Parametric modulation results for social reward prediction encoding in control participants (LD) only.*

	MNI coordinates				
Brain Region	X	Y	Z	Z score	*p* value
Right Inferior Temporal Lobe	52	-36	-22	4.40	0.025
Right Superior Temporal Lobe	44	-24	-4	3.21	
Right Fusiform Gyrus	38	-34	-22	3.12	

Whole-brain cluster p values are family-wise error corrected at $p < .05$

One Sample T-Tests

Visual inspection of the parameter estimates extracted from the peak voxels of the group contrast suggested that LD participants encoded social reward predictions positively, while HD participants appeared to encode them negatively (see Figure 5 in the main paper). To formally test this effect, one-sample t-tests against zero were performed separately for the two groups on the extracted parameter estimates. It was found that insula ($t(21) = 2.59$; $p = 0.017$) and parietal ($t(21) = 2.86$; $p = 0.009$) parameter estimates were significantly *above* zero in the

LD group, while they were significantly *below* zero in the HD group ($t(20) = 3.06$; $p = 0.006$; $t(20) = 3.06$; $p = 0.006$, respectively). This suggests that BOLD responses of LD individuals tracked the prediction value for happy faces, while neural responses of HD subjects appeared to track the prediction value for neutral faces.

This suggestion was further supported by whole-brain one sample t-tests, which revealed that HD subjects demonstrated *inverse* social reward prediction encoding in a parietal lobe cluster (MNI coordinates: 22 -64 56; $Z = 3.69$; $p_{uncorrected} = 0.003$; although this result did not quite reach significance after family wise error correction on the cluster level; $p_{FWE\text{-}corrected} = 0.192$). By contrast, LD participants did not show any encoding of inverse social reward prediction values (even at an uncorrected cluster level threshold). However, as reported in the main paper, LD subjects did display *positive* reward prediction encoding in the temporal lobe and fusiform gyrus, while no such effects were seen in HD individuals.

ROI Analysis

A recent meta-analysis identified the subgenual anterior cingulate cortex (sgACC) as the only region which consistently encoded model-derived prediction values across studies (Chase et al., 2015). Thus, a region of interest analysis was performed on this area. For this purpose, MarsBar (Brett et al., 2002) was used to extract prediction-related parameter estimates from a 8mm sphere (as in Ham, Greenberg, Chase, & Phillips, 2016) around the sgACC coordinates indicated in the meta-analysis (ROI 1: 4 34 -6; ROI 2: -6 28 -20). A one-way ANOVAs performed on the extracted parameter estimates revealed no group differences for social reward prediction (ROI 1: $F(1,41) = 0.01$, $p = 0.932$; ROI 2: $F(1,41) = 0.37$, $p = 0.545$) or social aversion prediction (ROI 1: $F(1,41) = 2.56$, $p = 0.117$; ROI 2: $F(1,41) = 1.22$, $p = 0.276$) encoding.

Analysis with Individual Parameters

When individual parameter values were used in the computational model to derive prediction values for the parametric modulation analysis, similar results were obtained as when average parameters were used (as in section 3.2.1 of the main paper). Specifically, it was found that HD subjects showed reduced social reward prediction encoding in the precuneus, inferior parietal lobe and superior temporal lobe compared to LD controls (see Table S2). No significant group differences were observed for social aversion prediction encoding.

Table S2: *Parametric modulation results for social reward prediction encoding in individuals with low (LD) vs high (HD) depression scores using individual modelling parameters*

	MNI coordinates				
Brain Region	X	Y	Z	Z score	*p* value
LD > HD					
Precuneus	20	-50	46	3.18	0.005
Inferior Parietal Lobe	32	-58	48	3.12	
Superior Temporal Lobe	38	-56	18	3.26	0.001

Whole-brain cluster p values family-wise error corrected at $p < .05$

Neural Responses to Name Cues and Emotional Faces

None of the name cue or face contrasts resulted in any significant clusters in the LD group alone. Yet, group comparisons revealed significantly higher activation to fearful (vs. neutral) faces in HD compared to LD subjects in the bilateral supramarginal gyrus, right fusiform gyrus, bilateral inferior temporal lobe, dorsal anterior cingulate, and in a cluster ranging from the dorsolateral to the ventrolateral PFC and to the insula (see Table S3). No group differences were observed for the happy vs. neutral face contrast or for the social reward vs. social aversion name cue contrast.

Table S3: *Regions showing higher responses to fearful (vs. neutral) faces in individuals with high (HD) compared to low (LD) depression scores*

	MNI coordinates				
Brain Region	X	Y	Z	Z score	*p* value
HD > LD					
Dorsal ACC/ MCC	-2	10	28	4.73	<0.001
Right Occipital Lobe	18	-92	-8	4.30	0.033
Right Fusiform Gyrus	34	-76	-18	3.56	
Right dlPFC (BA 8)	50	24	42	4.25	<0.001
Right vlPFC (BA 45)	54	32	10	3.50	
Right Insula	46	10	12	3.18	
Right Supramarginal Gyrus	36	-46	50	4.01	<0.001
Right Inferior Temporal Lobe	58	-54	-4	3.99	0.002
Left Inferior Temporal Lobe	-54	-58	-14	3.96	0.034
Left Supramarginal Gyrus	-28	-48	52	3.36	0.001

Whole-brain cluster p values family-wise error corrected at $p < .05$; ACC, anterior cingulate cortex; MCC, mid cingulate cortex; dlPFC, dorsolateral prefrontal cortex; vlPFC, ventrolateral prefrontal cortex; BA, Brodmann Area

References

Bakker, J.M., Goossens, L., Kumar, P., Lange, I.M.J., Michielse, S., Schruers, K., Bastiaansen, J.A., Lieverse, R., Marcelis, M., Amelsvoort, T. Van, 2018. From laboratory to life : associating brain reward processing with real-life motivated behaviour and symptoms of depression in non-help-seeking young adults. Psychol. Med. https://doi.org/10.1017/S0033291718003446

Behrens, T.E.J., Hunt, L.T., Woolrich, M.W., Rushworth, M.F.S., 2009. Associative learning of

social value. Nature 456, 245–249. https://doi.org/10.1038/nature07538.Associative

Brett, M., Jean-Luc, A., Valabregue, R., Poline, J.-B., 2002. Region of interest analysis using an SPM toolbox. Present. 8th Int. Conf. Funct. Mapp. Hum. Brain Sendai, Japan. Available on CD-ROM in NeuroImage.

Chase, H.W., Kumar, P., Eickhoff, S.B., Dombrovski, A.Y., 2015. Reinforcement learning models and their neural correlates: An activation likelihood estimation meta-analysis. Cogn. Affect. Behav. Neurosci. https://doi.org/10.3758/s13415-015-0338-7

Chowdhury, R., Guitart-Masip, M., Lambert, C., Dayan, P., Huys, Q., Düzel, E., Dolan, R.J., 2013. Dopamine restores reward prediction errors in old age. Nat. Neurosci. https://doi.org/10.1038/nn.3364

Collins, A.G.E., Frank, M.J., 2012. How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. Eur. J. Neurosci. 35, 1024–1035. https://doi.org/10.1111/j.1460-9568.2011.07980.x

Daw, N.D., 2011. Trial-by-trial data analysis using computational models, in: Delgado, M.R., Phelps, E.A., Robbins, T.W. (Eds.), Decision Making, Affect, and Learning: Attention and Performance XXIII. Oxford University Press, Oxford, pp. 3–38. https://doi.org/10.1093/acprof:oso/9780199600434.003.0001

Daw, N.D., O'Doherty, J.P., Dayan, P., Seymour, B., Dolan, R.J., 2006. Cortical substrates for exploratory decisions in humans. Nature. https://doi.org/10.1038/nature04766

Ekman, P., Friesen, W.V., 1976. Pictures of Facial Affect. Consulting Psychologists Press, Palo Alto. https://doi.org/citeulike-article-id:4270156

Gradin, V.B., Kumar, P., Waiter, G., Ahearn, T., Stickle, C., Milders, M., Reid, I., Hall, J., Steele, J.D., 2011. Expected value and prediction error abnormalities in depression and schizophrenia. Brain 134, 1751–1764. https://doi.org/10.1093/brain/awr059

Ham, B.J., Greenberg, T., Chase, H.W., Phillips, M.L., 2016. Impact of the glucocorticoid

receptor Bcl i polymorphism on reward expectancy and prediction error related ventral striatal reactivity in depressed and healthy individuals. J. Psychopharmacol. 30, 48–55. https://doi.org/10.1177/0269881115602486

Hindi Attar, C., Finckh, B., Büchel, C., 2012. The influence of serotonin on fear learning. PLoS One 7. https://doi.org/10.1371/journal.pone.0042397

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R.J., Frith, C.D., 2006. Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. Nature 442, 1042–1045. https://doi.org/10.1038/nature05051

Rescorla, R., Wagner, A., 1972. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement, in: Classical Conditioning: Current Research and Theory, Vol. 2. https://doi.org/10.1101/gr.110528.110

Rothkirch, M., Tonn, J., Köhler, S., Sterzer, P., 2017. Neural mechanisms of reinforcement learning in unmedicated patients with major depressive disorder. Brain 140, 1147–1157. https://doi.org/10.1093/brain/awx025

Rutledge, R.B., Moutoussis, M., Smittenaar, P., Zeidman, P., Taylor, T., Hrynkiewicz, L., Lam, J., Skandali, N., Siegel, J.Z., Ousdal, O.T., Prabhu, G., Dayan, P., Fonagy, P., Dolan, R.J., 2017. Association of neural and emotional impacts of reward prediction errors with major depression. JAMA Psychiatry 74, 790–797. https://doi.org/10.1001/jamapsychiatry.2017.1713

Schonberg, T., Daw, N.D., Joel, D., O'Doherty, J.P., 2007. Reinforcement Learning Signals in the Human Striatum Distinguish Learners from Nonlearners during Reward-Based Decision Making. J. Neurosci. https://doi.org/10.1523/JNEUROSCI.2496-07.2007

Schonberg, T., O'Doherty, J.P., Joel, D., Inzelberg, R., Segev, Y., Daw, N.D., 2010. Selective impairment of prediction error signaling in human dorsolateral but not ventral striatum in Parkinson's disease patients: evidence from a model-based fMRI study. Neuroimage 49,

772–781. https://doi.org/10.1016/j.neuroimage.2009.08.011

Tobia, M.J., Guo, R., Schwarze, U., Boehmer, W., Gläscher, J., Finckh, B., Marschner, A., Büchel, C., Obermayer, K., Sommer, T., 2014. Neural systems for choice and valuation with counterfactual learning signals. Neuroimage 89, 57–69. https://doi.org/10.1016/j.neuroimage.2013.11.051

Wilson, R.C., Niv, Y., 2015. Is Model Fitting Necessary for Model-Based fMRI? PLoS Comput. Biol. https://doi.org/10.1371/journal.pcbi.1004237